

不完全情報型ゲーム”ぷよぷよ”の勝率向上アルゴリズム

上田圭造^{†1} 梶並知記^{†2} 鈴木孝幸^{†1} 清原良三^{†1}

概要: 不完全情報ゲームである“ぷよぷよ”は“ぷよ”と呼ばれるブロックが落ちるゲームの1種であり、2個先まで落ちる“ぷよ”の配色はわかるが、それ以降はランダムである。コンピュータとの対戦に関して、その手法の提案がいくつか存在するものの、高度な人間のプレイヤーには勝てない場合も多い。現状のコンピュータのあるリズムが、大きな“ぷよ”の連鎖を作成することに主眼を置いている場合が多い。たしかに大きな連鎖ができれば効果絶大であるものの連鎖の処理に時間を要するなど、小さな連鎖と比較した場合に必ずしも有利とは言えない。そこで、小さな連鎖を狙うようなアルゴリズムを提案し、既存のアルゴリズムと対戦させ、その強さを評価したので報告する。

キーワード: ぷよぷよ, 連鎖, コンピュータゲーム

1. はじめに

近年,ゲーム AI は急速に発達し,チェス,将棋などの完全情報ゲームでは,AI の強さは人間のトッププレイヤーを上回るようになった。チェスでは 1997 年にグランドチャンピオンを,将棋では 2013 年にプロ棋士を,囲碁では 2015 年にプロをそれぞれ負かしている。

また,存在する不完全情報ゲームの AI の例としてテトリスと麻雀がある。2019 年にはテトリスの AI がトッププレイヤーに勝利を収めているが,その勝率は完全情報ゲームの AI に比べて低い。同年に麻雀の AI では天鳳の最高位の次点にあたる 10 段を達成したが,実力を示す指標である安定段位は 8.7 段となっている。安定段位が 8.7 段以上のプレイヤーは一定数おり,それらのプレイヤーと比べると劣っている。

これらの不完全情報ゲームでは,未だ AI の強さはそれらのプレイヤーと比べると劣っている。その原因の一つに,ゲームの状態の特徴量の取得やルールの明示化が難しいことが挙げられる。また,確率的な要素も多く,コンピュータでの探索時間も限られるという点からも AI というより,単純なプログラムでもあることも多いためと考える。本稿では,コンピュータ側のプログラムを“COM”と呼ぶこととする。

不完全情報ゲームでも,麻雀では,場に出た牌と自らの牌の情報が見えているため,隠れている牌や山に残っている牌の確率が計算できる即ち常に条件付き確率となる特徴があり,手が進むにつれて推測が容易になると考えられるが,テトリスや“ぷよぷよ”では落下する塊は完全にランダムであり,数に制限がないため,どの塊が出るかは常に独立の事象となる。

また“ぷよぷよ”では,“おじゃまぷよ”というものがあ

り,対戦相手に影響を与えることができる。この要素があるため,最適な戦略を単純に計算できなくなるという特徴がある。

そこで,本研究では,不完全情報の落下型パズルゲーム“ぷよぷよ”において,直接的に示されていない情報を予測し,最適な戦術を実現することが目的とし,“ぷよぷよ”の状態特徴量を用いた戦術的アルゴリズムを提案する。

2. “ぷよぷよ”とは

“ぷよぷよ”は,株式会社コンパイルが 1991 年に開発した落下型パズルゲームの一種である。落下してくる 2 つ 1 組の“ぷよ”を操作して大連鎖を構築するが,3 手以上先の“ぷよ”の配色を事前に知ることができない。このように場に晒されていない情報が存在するため,多くの可能性を考慮する必要があり,先読みが難しい。

以下にぷよぷよのルールの説明をする。

1. 図 1 に示すようにフィールドは縦 12 × 横 6 マスの格子で構成され,格子の 1 マスにつき 1 つの石(以下,“ぷよ”と呼ぶ)を置くことができる。
2. “ぷよ”には 4 つの色があり,フィールド上部からランダムな配色の“ぷよ”が 2 つ 1 組で落下する。プレイヤーはその“ぷよ”に対して,回転,横移動,高速落下のいずれかの操作を行う。落下する“ぷよ”はフィールドの枠外に「NEXT”ぷよ””として 2 手先まで表示される。
3. 落下した“ぷよ”が床や下にある“ぷよ”に衝突すると,その位置に“ぷよ”が固定される。片方の“ぷよ”だけ衝突した場合,もう片方の“ぷよ”は強制的に衝

^{†1} 神奈川工科大学
Kanagawa Institute of Technology

^{†2} 文教大学
Bunkyo University

突するまで落下する。これをちぎりと呼ぶ。“ぷよ”をちぎって設置するともう片方の“ぷよ”にちぎりモーションが入り、設置時間が延長される。ちぎるによる段差が大きいほど設置時間は長くなる。

4. 図2に示すように縦横に同色の“ぷよ”が置かれたとき、それらは結合する。4つ以上結合した場合に消滅し、得点となる。
5. “ぷよ”の消滅により、上にあった“ぷよ”は落下する。このとき再び“ぷよ”が4つ以上結合すると消滅し、連鎖が起きる。連鎖数が多いほど得点は高くなる。
6. 得点に応じて、相手のフィールドの上部に“予告ぷよ”が表示される。“予告ぷよ”は相手プレイヤーが“ぷよ”を落とした直後、“おじゃまぷよ”としてフィールドに落下する。この“おじゃまぷよ”は上下左右に隣接する“ぷよ”が消えると同時に消える特性を持つ。
7. 自分のフィールドに“予告ぷよ”が表示されている場合に連鎖を行うと、“予告ぷよ”の量を減らすができる。これを相殺という。また、“予告ぷよ”がある状態で相手の得点より大きな連鎖を行うと相手との得点の差が相手のフィールドの上部に“予告ぷよ”として表示される。
8. 左から3列目の12段目が“ぷよ”で埋まるとゲームオーバーとなる。

本論文では“ぷよぷよ”の操作について、以下のように定義する。

- 連鎖を開始することを発火と呼ぶ。
- “ぷよ”の操作を終え、次の“ぷよ”を操作することを次の“ぷよ”を引いてくるという。

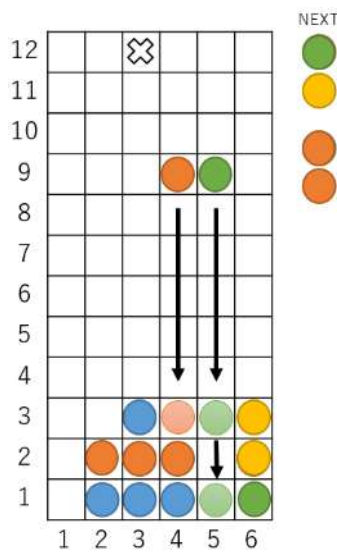


図1 ぷよの落下状態

3. 関連研究

3.1 対戦型ゲームにおける戦略推定に関する研究

対戦型ゲームにおいて戦略を推定する研究は、ゲーム開発におけるゲームバランスの調整[1]や、人間プレイヤーの実力と互角になるように調整したゲーム接待 AI の提案[2]に利用されている。

文献[1]ではゲームバランスを「戦略の優位性の公平さ」、戦略を「利得獲得を目的としたプレイヤーの行動の時系列」とそれぞれ定義している。そして他プレイヤーの行動に関わらず、プレイヤーにとって他のすべての戦略よりも大きい利得が得られるような戦略を支配的戦略と呼んでおり、これが存在する場合はゲームバランスが悪いと評価している。支配的戦略を探索し、ゲームバランスの分析、可視化を自動的に行う手法を提案することでゲームバランスの調整にかかる時間的コストの削減につながる。ぷよぷよはプレイヤー同士の対戦における条件(各プレイヤーが引いてくる“ぷよ”

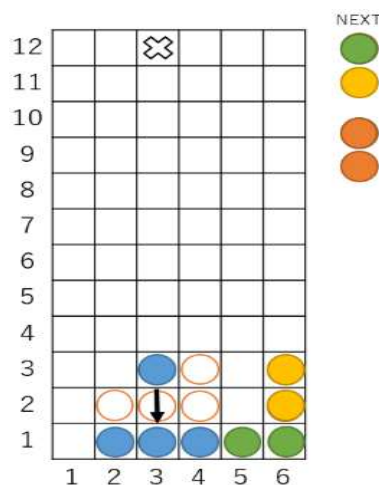
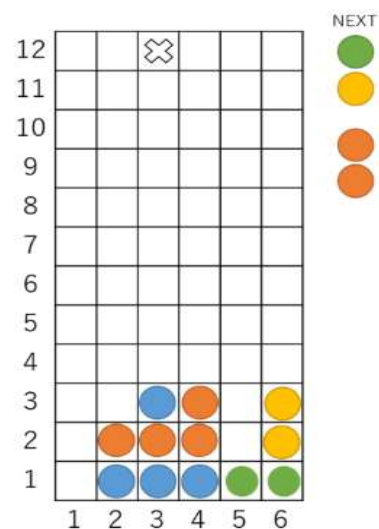


図2 連鎖の例

の色の組合せと順番はそれぞれ等しい)は同じであるため、支配的戦略が発生することはない。よって良いゲームバランスを持ったゲームであると言える。

文献[2]ではCOMを対戦相手とした2人対戦型ゲームにおいてゲーム固有の知識を用いることなく一般にプレイヤーの強さを推定し、それに応じた強さでありながら不自然さを感じさせないようなCOMを用意する手法を作成し、プレイヤーがCOMを相手とした対戦ゲームにおいてより楽しさを感じるようにすることを目標としている。また、最善の戦略が存在せず、戦略同士に相性があるような複雑な対戦型ゲームにおいて、COMの調整にはプレイヤーの戦略を含めたプレイ内容の推定が必要だと考えられる。本研究は人間のプレイヤーを対象とした戦略ではなく、COMを対象にし、その性能向上を目的とした戦略の提案である。

3.2 連鎖作成時の自動探索に関する研究

“ぷよぷよ”のCOMのアルゴリズムは多数提案されている。例えば、文献[3]は、関連性行列を用いた定型連鎖構成に関しての手法を示している。この手法を利用して実装された方式は、連鎖構築に関して人間のトッププレイヤーに匹敵する。しかし、連鎖を組み終わる前の中盤に相手に、“おじゃまぷよ”を送られると連鎖を作るフィールドが狭くなり、敗北につながる。

これを解消するには自分から相手に攻撃を仕掛けるか、相手を迎撃する必要がある。そこで本論文では中盤以降に副砲といわれる連鎖を常に保持するアルゴリズムを提案する。

4. 既存COM(シミュレータ)

4.1 動作環境と操作時間

既存のCOM[4]の実行時間などを調査した。将棋や囲碁と違い、“ぷよぷよ”では瞬時に次の手を決める必要があり、時間のかかる探索などはできない。今回、アルゴリズムを実装した環境を表1に示す。

4.2 時間と戦術

本節では、ぷよぷよにおける基本的な戦術を説明する。ぷよぷよの対戦における勝利条件は相手のフィールドの左から、3列12段目を埋めることである。相手のフィールドを埋めるには連鎖を発火し、“おじゃまぷよ”を相手にフィールドに送る必要がある。“ぷよぷよ”における相殺というシステムは、相手と自分の連鎖の大きさの差分を“おじゃまぷよ”として送る。よって、一般に相手より大きな連鎖を作る必要がある。今回使用したシミュレーションソフトウェアでのぷよの各操作と連鎖時にかかる時間を表2に示す。この表からわかるように連鎖は発火してから次のぷよが操作可能になるまで、インターバルが生じる。このインターバルは連鎖数が多いほど長くなるため、この間に新しい“ぷよ”を引いてきて相手より1連鎖大きくすれば有利状況を作ることができる。

表1 実行環境

CPU	Intel Core i7-7500U 2.70GHz
OS	Windows 10
メモリ	8GB

表2 ぷよの操作にかかる時間(秒)

ぷよを1マス下へ移動	4/60
ぷよを1マス横へ移動	8/60
ぷよを最上段から最下段へ移動	44/60
ぷよを中央から最右列へ移動	24/60
ちぎりモーション中に1マス下へ落下	14/60
ぷよを設置後、連鎖によるぷよ消滅までにかかる時間(ぷよの設置時間)	134/60
連鎖によるぷよ消滅後、次の連鎖のぷよ消滅にかかる時間	144/60
連鎖終了時もしくはぷよを設置時から次のぷよを操作できるまでの時間	66/60

例えば、図3では左のプレイヤーが3連鎖を発火している。この場合、表2より連鎖終了までにかかる時間は $134+144 \times 2+66=488$ (/60秒)となる。この時、右のプレイヤーはNEXTの青緑のぷよを1列目に縦に置き、その次の黄のぷよを5列目に置けば4連鎖を発火でき、有利な状況を作ることができる。黄のぷよを引くことができれば、その操作が終わるまでおじゃまぷよは降らないため、かかる時間は黄のぷよを引くまでの $4 \times 9+8 \times 2+66=118$ (/60秒)となる。

このように相手プレイヤーより後に連鎖を発火し、有利状況を目指すことがぷよぷよにおける基本的な戦術となる。

4.3 既存のCOMのアルゴリズム

本節では、既存COMにおける連鎖構築の評価方法について述べる。

(1) 連鎖の評価

連鎖は以下の三点を評価している。

- 連鎖数
連鎖の得点が大きいほど高い評価値を与える。
- 連鎖の形
連鎖構築の際に段差による凹凸が激しい場合、ちぎりの頻度が多くなるため、減点する。
- 連鎖までに必要な“ぷよ”の数
連鎖を発火するまでに必要なぷよの数を引くための時間の期待値を計算する。“ぷよ”は4色の中からランダムに2つ選ばれる。よって、必要な“ぷよ”が多ければ

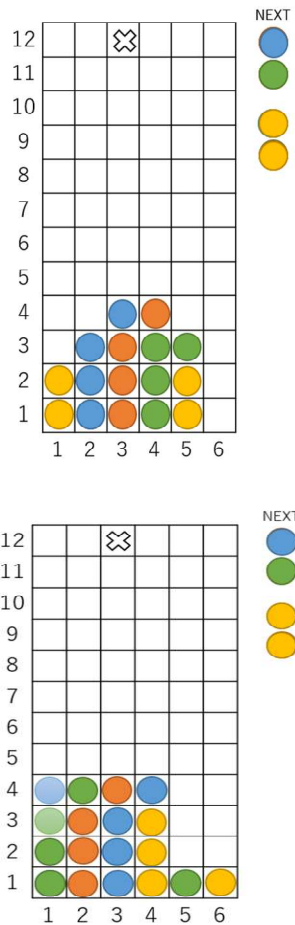


図3 ぶよぶよの基本戦術

より連鎖開始に時間がかかる確率が高くなり、減点となる。

(2) 相互連鎖結果予測アルゴリズム

自分と相手の想定される連鎖の選択肢を列挙し、お互いの連鎖をシミュレーションすることで送られる“おじゃまぶよ”の量を予測する。このアルゴリズムでは、現在発火することができる連鎖を親ノード、その連鎖を発火した後にできる連鎖を子ノードとした木構造で表現している。また、各連鎖のノードにはその連鎖を打つまでに必要な時間と連鎖が終了するまでの時間の情報が含まれており、自分と相手の状況の把握と連鎖を発火するタイミングを判断する役割がある。

現状、既存のアルゴリズムでは自分から連鎖を発火し、仕掛けることは少ない。相手の連鎖を確認してから対応した方がより堅実なプレイができるためである。しかし、自分から仕掛けた方が有利状況を作れる場面で、連鎖を仕掛けないことは勝率向上の面においてボトルネックとなってい

る。本稿では、連鎖実行時間の短い連鎖を保持し、相手に隙がある状態で、自分から保持している連鎖を仕掛けるアルゴリズムを提案することでこの課題点の解決を試みた。

5. 提案手法

5.1 アルゴリズム

本節では、提案するアルゴリズムの概要について述べる。本アルゴリズムのフローチャートを図4に示す。

アルゴリズムの概要は、お互いが一定の連鎖数以上を保持している場合に、連鎖実行時間の短い連鎖を作成・保持し、相手の連鎖作成時に一定の隙があるタイミングで、保持していた連鎖を発火するというものである。これにより、相手フィールドに“おじゃまぶよ”を送り、有利状況を作ることが期待される。

一定の連鎖数のある連鎖を保持しているか否かを判断する条件は、対戦における中盤に差し掛かっているか否かの判断に必要である。今回の提案手法では5連鎖以上を保持している場合を中盤と定義する。4.2節で説明した戦術を行うために、中盤以降は如何に相手に先に大きな連鎖を発火させるかという駆け引きをすることになる。

また、一定の隙とは、「任意の連鎖を開始するために必要な個数の“ぶよ”を引いてくるまでの時間のことである。“ぶよ”は4色の中からランダムに2つ選ばれるため、相手の連鎖開始までにかかる時間は予想値となる。

5.2 アルゴリズムの適用条件

提案するアルゴリズムの適用条件として、アルゴリズムによって作成した短い連鎖を発火してから連鎖終了までにかかる時間が、「相手が任意の連鎖を開始するために必要な個数の“ぶよ”を引いてくるまでの時間より短い必要がある。よってあまりにも作成した短い連鎖の連鎖実行時間が長いと連鎖を発火する条件を満たさない場合がある。そのため、短い連鎖の連鎖実行時間の上限を、2連鎖を発火した際に生じる連鎖実行時間と等しい5.7秒と定義する。

6. 評価

既存のCOMに提案手法を適用したCOMと従来手法のCOMを対戦させ、アルゴリズム適用後の各対戦場面の有利・不利状況の検証を行った。

有利・不利状況の判断は、相手に送った“おじゃまぶよ”によって相手の保持している連鎖の選択肢の変化をもとに判断をする。例えば、図5では黒の“ぶよ”を“おじゃまぶよ”として表しているが、“おじゃまぶよ”がない場合、連鎖の選択肢は黄→緑の2連鎖、赤→青の2連鎖、緑→青の2連鎖の3つがある。しかし、“おじゃまぶよ”が図3のようにある場合、連鎖の選択肢は緑→青の2連鎖のみになる。この時、連鎖の選択肢は減り、行動が制限されるため、“おじゃまぶよ”がない場合に比べ、不利になったと言える。

アルゴリズム適用後の各対戦場面の有利・不利状況を

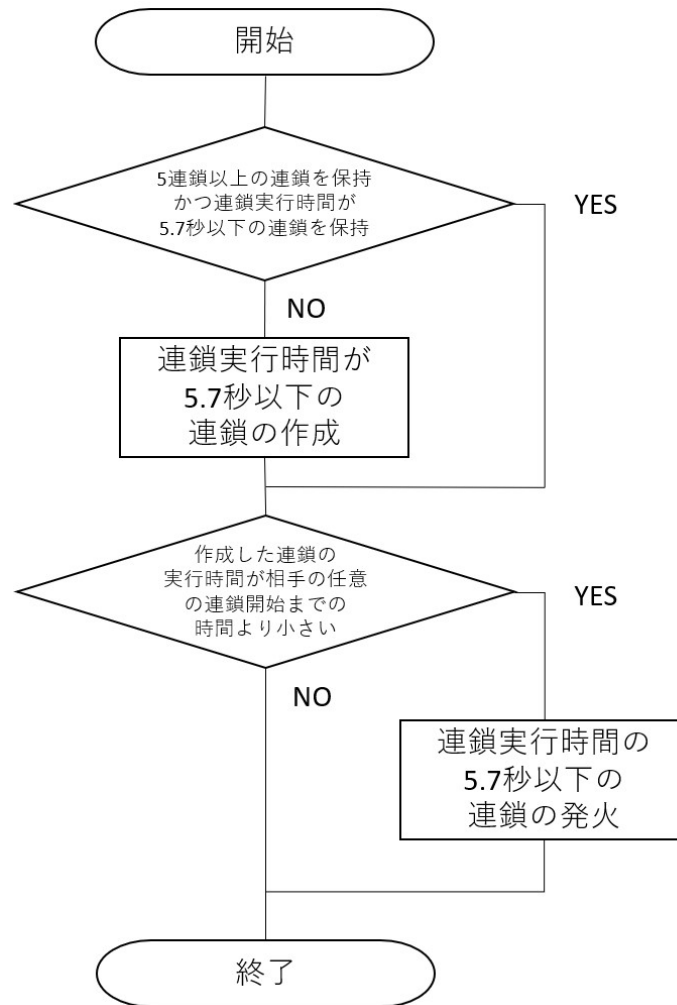


図4 アルゴリズム概要

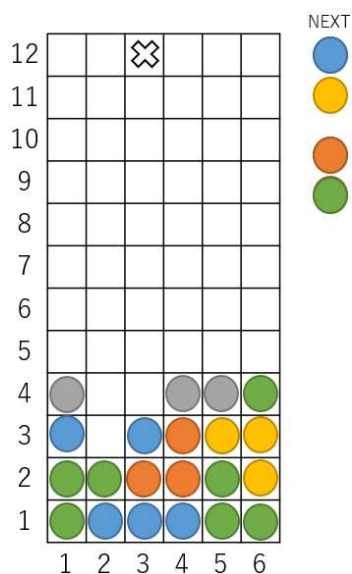


図5 連鎖の選択肢

“おじゃまふよ”が降る前と後で分析した結果を表3に示す。

この結果からわかるように相手に“おじゃまふよ”を降らせ、相手の保持している連鎖の選択肢を減らすことに成功し、有利状況を作ることができた。

しかし、アルゴリズムが適用される状況は限定的であり、得られたデータ数は少ない。原因としてアルゴリズムの適用条件を達成する頻度が少ないことが挙げられる。また、相手の連鎖の形によっては連鎖の選択肢を減らすことができず、ただ時間と連鎖を無駄にしてしまうことがあった。

7. おわりに

本論文では、“ふよふよ”の対戦における戦略的アルゴリズムを提案した。既存COMでは対戦の中盤において自分から連鎖を発火することは少なく、勝率向上の妨げになっていた。このため、著者らは中盤以降に実行時間の短い連鎖を保持し、隙を窺って発火して有利状況を作り出すことで課

表3 おじゃまぶよの有無による連鎖の選択肢の変化

	おじゃまぶよの数	降る前	降った後
1	8	2	2
2	24	3	0
3	16	3	0
4	10	2	1
5	2	3	2

題点の解決にあたった。しかし、アルゴリズムが適用される状況は限定的であり、得られたデータ数は少ない。また、実行時間の短い連鎖を作成・保持した後に発火する機会がなく対戦が終わってしまう場面も多々あった。アルゴリズムの適用条件を調整し、再度実験を行い確かめる必要がある。

参考文献

- [1] 山本界人, 水野竣介, ターウォンマット ラック: 対戦型ゲームにおける戦略多様性についての StGA 法を用いた自動分析手法の提案とその評価, 情報処理学会論文誌, Vol.55, No.11, pp.2328-2335 (Nov. 2014)
- [2] 杉本直樹, 鶴岡慶雅: 戦略の動的推定による 2 人対戦ゲーム接待 AI の提案, The 23rd Game Programming Workshop 2018, pp.114-119
- [3] 富沢大介, 池田心: 落下型パズルゲームの定石形配置法とぶよぶよへの適用, 情報処理学会論文誌, Vol.53, No.11, pp.2560-2570 (2012)
- [4] “GitHub - puyoai/puyoai: AI for puyo”.
<https://github.com/puyoai/puyoai>, (参照 2020-05-04).